Küssner, M. B. & Eerola, T. (2019). The Content and Functions of Vivid and Soothing Visual Imagery during Music Listening: Findings from a Survey Study. *Psychomusicology: Music, Mind, and Brain. Accepted manuscript (author version)/.*

Abstract

Studies have suggested that visual imagery forms an important part of the listening experience, and might be one of the mechanisms by which music induces emotions in a listener. However, little is known about the content, prevalence and functions of visual imagery during music listening. To that end, an online survey was constructed to explore music-related visual imagery. This included 24 statements about the visual imagery based on prior research and an open question about the content of the inner images. Several standardized questionnaires (VVIQ, Gold-MSI) were included as well to investigate the link to visual imagery in general and musical training. In total, 669 participants provided responses to an online survey. A factorial structure of music and visual imagery statements provided a 3-factor structure consisting of vivid, soothing and disruptive visual imagery, although the actual factor structure was non-identical between the musically trained and untrained respondents. Separate analyses of factor for musically trained and untrained participants yielded a more parsimonious structure of visual imagery, which consisted of vivid and soothing visual imagery. These two factors exhibited consistently different weights across the items; for musically trained participants, the vivid imagery was more related to modulating the arousal. The ability to conjure up vivid visual imagery was only weakly related to the music-related visual imagery. A content analysis of the open question revealed common themes that related to a mixture of concrete visual imagery (landscapes, images of people, scenes from past events) and abstract visual imagery (shapes, objects and colours). Implications of these findings for further studies on music-induced emotions are discussed with a focus on a recent constructionist account of emotional meanings in music.

Keywords: music listening, visual imagery, emotion, musical training, online survey

The Content and Functions of Vivid and Soothing Visual Imagery during Music Listening: Findings from a Survey Study

Music is present in all cultures, and most people experience moments in which they are moved by music. They can experience powerful emotional responses to music triggered by a song of their favourite band, a special performance or occasion, a personal memory associated with a particular piece, or images they conjure up during listening. For instance, individuals may visualize internal images that consist of pictorial representations (e.g., landscapes, people, past events) or embodied image-schemata (e.g., images of ascending or descending motion). Juslin and Västfjäll (2008) propose that visual imagery is one of six mechanisms by which music induces emotions in the listener, the other mechanisms being brain stem reflex, evaluative conditioning, emotional contagion, episodic memory, and musical expectancy; the latest version of this framework also includes rhythmic entrainment and aesthetic judgment (Juslin, 2013). Juslin (2013) suggests that visual imagery in general develops during pre-school years and is strongly influenced by culture and learning. Unlike other emotion-induction mechanisms such as the brain stem reflex or evaluative conditioning which are automatic and subliminal, visual imagery in response to music is typically experienced consciously and therefore supposed to be under control by the listener, i.e. visual imagery can be changed or suppressed by the individual. Before reviewing the relatively sparse literature on music and visual imagery, it is important to note that music is unambiguously associable with vision.

Research on cross-modal mappings of music (Eitan, 2017), on graphical representations of music (Küssner, 2013; Tan & Kelly, 2004) and on audio-visual correspondences (Küssner & Leech-Wilkinson, 2014; Spence, 2011) have significantly advanced our understanding of how auditory and visual senses interact. Further evidence for the association between music and vision can be found in the substantial literature on music in film (Bullerjahn, 2016; Cohen, 2011; Kuchinke, Kappelhoff, & Koelsch, 2013). Particularly the effect of the visual (usually narrative) on the experience of music has been shown in several studies (Boltz, Ebendorf, & Field, 2009; Goldberg, Chattopadhyay, Gorn, & Rosenblatt, 1993), and there is strong evidence that a combination of music and image elicits emotions (Eldar, Ganor, Admon, Bleich, & Hendler, 2007; Geringer, Cassidy, & Byo, 1997; Vines, Krumhansl, Wanderley, & Levitin, 2006). Remarkably, Eldar and colleagues (2007) were able to show that combining music and film elicited increased activity in brain regions associated with emotional processing whereas music alone did not evoke a differential response in these areas.

The potential of visuals—or indeed, visual imagery—in combination with music has also been exploited in therapeutic contexts. Most research has been carried out in the realm of music therapy (Band, Quilter, & Miller, 2001; Bonde, 2005), involving the development of the “Bonny Method of Guided Imagery and Music” which originated in the 1970s (Bonny, 2002) and may be used to treat affective disorders such as anxiety, depression or stress. This method, which traditionally includes five phases in which a therapist guides the visual imagery (prelude, induction/relaxation, music evoked imagery with verbal dialogue, transition with mandala drawing, and postlude), is one of most advanced models of receptive music therapy today (Bonde, 2015). The efficacy of music and visual imagery has also been demonstrated in other clinical contexts, e.g. with cancer patients. Karagozoglu, Tekyasar, and Yilmaz (2013) showed that combining music therapy and visual imagery significantly reduced side effects of chemotherapy such anxiety, nausea and vomiting.

While visual imagery on its own has been studied widely in psychology (Broggin, Savazzi, & Marzi, 2012; Kolers, 1983) and cognitive neuroscience (Kosslyn et al., 1999), there is relatively little research concerning the relationship of music and visual imagery. Using content analysis in an exploratory study, Osborne (1981) showed that visual imagery was the most frequently reported response upon listening to synthesized electronic music among four identified categories including thoughts, emotions and bodily sensations. Further evidence that music facilitates visual imagery is provided by Quittner and Glueckauf (1983) who demonstrate that listening to music produces significantly more visual imagery than a control or a relaxation condition. There is also empirical support for inter-individual differences in the ability to conjure up images during music listening. Stratton and Zalanowski (1992) probed the amount of suggested visual imagery in three conditions: matching music, non-matching music and silence. They found that a mismatch between the title of classical music excerpts and verbal descriptions of scenes led to a decrease in visual imagery, while no difference between the conditions silence and matching music was observed. However, it was revealed that individuals' amount of right hemisphere style (assessed with questionnaires) correlated positively with vividness of visual imagery in the no-music condition, but no correlation was present in the matching music condition, suggesting an interference of these two tasks—music listening and visual imagery—in the right hemisphere. Specialized training might also affect visual imagery. Brochard and colleagues (2004) showed that musically trained participants exhibit enhanced performance (assessed with reaction time) in a visual perception and imagery task compared to musically untrained participants, which could partly be explained by better sensorimotor integration in musicians. However, specific expertise in a perceptual domain does not necessarily lead to enhanced vividness of visual imagery in that domain. Sunday, McGugin and Gauthier (2017) showed that domain-specific imagery (tested with car experts) correlates with general vividness of imagery, but not with perceptual or semantic expertise. Whether the same lack of domain-specific imagery abilities would be found in music experts needs to be tested.

To establish a more solid empirical foundation for such studies, we first need to investigate the exact nature of visual imagery during music listening since no one has attempted to systematically explore the contents of music-related visual imagery. Most previous research has focused on its effect on eliciting emotions during music listening. This has been demonstrated in a study in which participants were contacted at random times over the course of two weeks to report whether they had experienced a music-induced emotion and what its cause had been (Juslin, Liljeström, Västfjäll, Barradas, & Silva, 2008). While emotional contagion (i.e. feeling the same emotion that the music expresses) was the most commonly reported cause, visual imagery also came relatively high in the ranking – considerably ahead of musical expectancy. In a more recent study investigating the role of contextual information in music-induced emotions, Vuoskoski and Eerola (2015) found that up to 80% of their participants reported visual imagery when music was combined with narrative descriptions. However, the focus of this study was not on visual imagery; the findings emerged from open comments collected after a music-listening experiment and are part of a summary of broader themes related to listening experiences. In a similar vein, a potential link between visual imagery and emotional responses during music listening was reported in a neuroscientific study investigating neural correlates of music-induced joy and fear. Koelsch and colleagues (2013) discovered that listening to fear-inducing music was associated with an increased functional connectivity between the superficial amygdala and the visual cortex, hinting at the possibility that visual imagery may play a crucial role for the aesthetic appeal of fear-inducing music. Day and Thompson (same special issue) provide evidence that the relationship between visual imagery and felt emotions might be more complex. Participants were asked to indicate, in three separate counter-balanced conditions, (a) when they perceived an emotion in the music, (b) when they felt an emotional response themselves, and (c) when they experienced visual imagery. Results showed that it took significantly longer to experience visual imagery than to feel an emotional response. Thus, even though there is some empirical evidence for the association of visual imagery and music-induced emotions, exactly how this mechanism works is still poorly understood. A **more** fundamental problem is that we simply do not know what kind of inner images people experience during music listening, how common they are and to what extent visual imagery relates to more generic forms of mental imagery. These issues need to be clarified before any links to emotional responses can be meaningfully explored, because it is possible that emotions merely reflect the narrative of the visual imagery.

Aims. Our study on visual imagery during music listening therefore had three main goals:

1. Estimate the prevalence of visual imagery during music listening and collect detailed insights into the different kinds of visual imagery people experience while listening to music.
2. Explore how visual imagery in response to music is different from visual imagery in general.
3. Investigate how visual imagery correlates with domain-specific skills (Sunday et al., 2017).

Methods

Participants and procedure. Participants were obtained from two sources; 169 respondents were the result of a convenience sample recruited via professional mailing lists and social media (e.g., Twitter). A representative sample (N=500) of people living in the UK was obtained from Dalia Research (https://daliaresearch.com). We decided to pool the samples because we did not see any major differences in their background and for the purposes of the analyses. Both samples were collected with identical sets of questions, except for item #19, detailed later, that had a minor wording change[1] and one additional question about the prevalence of visual imagery during music listening[2]. The age range of the pooled sample was 18 to 79 years (M = 29.98 years, SD = 9.58 years); 381 (56.95%) respondents were female. Ethical approval for this study was obtained from the Ethics Committee of the Faculty of Humanities and Social Sciences at Humboldt University Berlin.

Questionnaire. The questionnaire consisted of existing instruments and 24 items designed to probe music-related visual imagery (MVI) developed based on past studies on visualizations of music (Küssner, 2013; Küssner & Leech-Wilkinson, 2014). These 24 statements are given in full in Appendix A. We also included the Vividness of Visual Imagery Questionnaire (VVIQ, Marks 1973), since that has been a successful self-report tool for capturing differences in people's ability to create visual imagery (Baltes & Miu, 2014; Campos, 2011). We also asked the respondents about their active and passive engagement with various forms of art using questions such as attendance to dance performances/plays or art exhibitions. To capture their musical engagement and sophistication, we utilized the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014) with its five subscales Active Musical Engagement, Self-reported Perceptual Abilities, Musical Training, Self-reported Singing Abilities, and Sophisticated Emotional Engagement with Music. We also asked them for basic demographic information (age, gender, education) and music preferences (using a list of genres). Finally, we provided an open-ended question about the content of music-related visual imagery (see Footnote 2).

[1] The wording of item #19 was changed from "The images I conjure up during music listening occur spontaneously" to "The images in my mind's eye during music listening occur spontaneously" to account for the fact that spontaneously occurring visual imagery does not involve the active process of conjuring up images.

[2] Since the prevalence could not be measured meaningfully with a convenience sample, we only included this question ("Have you ever experienced visual imagery (i.e. images in your mind's eye) while listening to music?") in the Dalia survey.

Results

Analysis strategy. The analysis will be broken down into four sections that explore the topic in an incremental fashion; the first will address the prevalence of visual imagery in music listening using the representative sample, the second will explore the structure of the visual imagery itself using the pooled sample of all participants and exploratory and confirmatory factor analyses, the third will scrutinize the impact of musical expertise on the discovered structures of visual imagery, and the last section will delve deeper into the contents of the music-related visual imagery by summarizing the qualitative responses provided by the participants to an open question about the contents of music-related visual imagery.

Prevalence. Visual imagery seems to be a common feature of music listening. Of 500 respondents of our representative sample, 77.20% (n=386) indicated that they have experienced visual imagery during music listening before, whereas 22.80% (n=114) reported that they have not experienced visual imagery.

Structure within the Music-Related Visual Imagery statements. To discover the possible structures within the music-related visual imagery, we will apply robust procedures for establishing the plausible structures within the 24 items representing MVI. Factor analysis is the most commonly used technique to discover the potential structure within correlated set of items. To maximise the robustness and generalisability of the structure we will discover, we first carry out Exploratory Factor Analysis (EFA) to identify the potential structure (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Subsequently, we will assess the robustness of the found solution with Confirmatory Factor Analysis (CFA) that computes the fit of the proposed structure derived from the EFA using a separate subset of the data (Schreiber, Nora, Stage, Barlow & King, 2006). To apply these analysis stages to our data, we randomly divide the full sample into training (60%) and test set (40%) and use the training set for the exploratory analysis and the test set for the confirmatory analysis.

In the initial phase, the factorability of responses across the statements was conducted using the Kaiser-Meyer-Olkin measure of sampling adequacy, yielding a robust indicator of factorability (0.86), although one item fell below the recommended threshold of 0.60 and was eliminated (#20 “The images I conjure up during music listening only last a few moments”). The optimal number of factors to extract was determined with several methods: Velicer’s MAP reduction algorithm (Zwick & Velicer, 1986), components from parallel analysis (Horn, 1965), and very simple structure (Revelle & Rocklin, 1979), as suggested by Ruscio and Roche (2012). This comparison of the optimal factors offered agreement among the methods for 3 components (as suggested by parallel and very simple structure). A factor analysis with oblimin rotation was utilised to increase the interpretability of the loadings. This model explained 37% of the variance and obtained a decent fit to the data (RMSEA = 0.074, RMSEA CI_{90}=0.066-0.077, *df*=187, X^2= 507.97). Items with factor loadings less than .45 and items that loaded on multiple factors were considered criteria for removal from the next round of analyses. The items removed were #4, #5, #6, #9, #13, #16, #19, and #21 (see Appendix A for the full set of MVI items).

The obtained factor structure was applied to the test set to estimate the fit of the model. In this CFA, we utilised the Lavaan package of R (version 0.5-23.1097) with robust maximum likelihood-based estimator, which corrects for non-normality, and used chi-square statistics and two recommended fit indices, root-mean-square error of approximation (RMSEA), and comparative fit index (CFI). According to Hu and Bentler (1999), the desirable cutoff value for the RMSEA should be .06, and for the CFI, it should be above .95. To improve the CFA model, we eliminated the items that achieved loadings <0.50 with the covariance matrix (Ximénez, 2009).

<INSERT TABLE 1 ABOUT HERE>

Table 1 displays the factor structure with the loadings from EFA and CFA. The first factor refers to active and vivid visual imagery (hereafter VIVID VISUAL IMAGERY) obtained with any recorded or live music (#1 and #2) and their reverse variant (#3). The imagery is abstract (#14) or dynamic (#17), it can happen with both eyes open (#23) and closed (#24). The second factor captures the emotional outcome of the process **(**labelled here as SOOTHING VISUAL IMAGERY**)**, namely whether visual imagery makes the participants feel relaxed (#10) or calm (#11). The third factor could be interpreted to be a factor where visual imagery is largely undesirable (hereafter DISRUPTIVE VISUAL IMAGERY), either the participant is trying to suppress the imagery (#7), the imagery bothers them (#15), and such imagery seems to be static as well (#16). The loadings presented in Table 1 are not especially high for either of the models (EFA or CFA), but the magnitude of the loadings across the training and test set is consistent (an average deviation of ± 0.085). The similarities between the models suggest that the CFA model is close to being acceptable, but strictly speaking the model fails to exceed the criteria for both measures of fit with CFI of 0.945 and RMSEA of 0.084 (see Table 2 for the full fit indices of the model).

<INSERT TABLE 2 ABOUT HERE>

Structure within participants – Musical training. In the first model, it could be that the structure is masked by interindividual differences. For instance, some people might find visual imagery as a problematic part of listening, as the factor 3 in the first analysis suggested, but others have a contrary experience and feel rather stimulated by it. Participants with musical training or those heavily engaged in music might actually want to experience visual imagery in conjunction with the listening, and hence factors 1 and 3 might inherently create conflicts across the responses given by participants with diverse interests and preferences or different levels of musical training. To explore this possibility more fully, ANOVA was applied to the factor scores from CFA where the impact of the Gold-MSI variables and the activity with respect to other arts (fiction, visual arts, dance, theatre) to the factor scores was examined. The scores of the first factor yielded a significant main effect of Training ($F(1,657)=33.63$, $p<.001$, $\eta^2_G=.003$), Perceptual ($F(1,657)=16.44$, $p<.001$, $\eta^2_G=.006$), Emotions ($F(1,657)= 29.64$, $p<.001$, $\eta^2_G=.039$), and engagement with fiction ($F(1,657)=9.87$, $p<.01$, $\eta^2_G=.002$). The second factor showed significant effects with Perceptual ($F(1,657)=14.10$, $p<.001$, $\eta^2_G=.005$), Training ($F(1,657)=37.67$, $p<.001$, $\eta^2_G=.005$, and Emotions ($F(1,657)=19.29$, $p<.001$, $\eta^2_G=.029$), and engagement with fiction, $F(1,657)=5.43$, $p<.05$, $\eta^2_G=.002$. The scores for the third factor showed a slightly similar pattern of main effects; Active ($F(1,657)=13.28$, $p<.001$, $\eta^2_G=.002$), Perceptual ($F(1,657)=59.49$, $p<.001$, $\eta^2_G=.015$), Singing ($F(1,657)=6.45$, $p<.001$, $\eta^2_G=.000$), Emotions ($F(1,657)=27.03$, $p<.001$, $\eta^2_G=.023$), and engagement with theatre ($F(1,657)=12.87$, $p<.001$, $\eta^2_G=.001$). These links between musical training and factor structures may suggest that the actual factors themselves might be slightly different depending on the level of musical training of the participants.

To address the dependence of factor structure on musical training, we applied cluster analysis to all normalised Gold-MSI variables. Two optimal clusters were identified by using a silhouette technique (Rousseeuw, 1987) that utilises a bootstrapping (1000 draws) to examine the clustering quality. The first cluster (n=450) contains the musically untrained participants (who score low on the Gold-MSI subscale Musical Training: M=16.81, SD=7.88) compared to the participants in the second cluster (n=219), who seem to be musically trained

(Gold-MSI score on Musical Training: M=33.01, SD=9.12; see Figure 1 for an illustration of the clusters and Gold-MSI variables).

<INSERT FIGURE 1 ABOUT HERE>

Having discovered the internal structure that divides the participants into two clusters, we can now re-estimate the factor solutions separately for both subsets, using similar 60%/40% training and testing subsets, which leaves us with 131 musically trained participants to build EFA model and 88 to test the model. For musically untrained participants, the EFA analysis suggests two factors, possibly labelled as VIVID VISUAL IMAGERY and SOOTHING VISUAL IMAGERY, where the vivid factor has items such as #18 “I see images in my mind’s eye whenever I listen to music”, #23 “I often conjure up images while listening to music with eyes open”, and #5 “The images I conjure up during music listening are one of the main reasons why I listen to music”. More importantly, this model provides a very good fit with the unseen data (RMSEA = 0.037, RMSEA CI_{90}=0.001-0.010, df=8, X^2= 10.01). When an identical analysis is conducted for musically trained participants, this yields also a two-structure solution, where both refer to emotions. The first one relates to vivid visual imagery (#5 “The images I conjure up during music listening are one of the main reasons why I listen to music”, #12 “The images I see in my mind’s eye when listening to music make me feel excited”, #13 “The images I see in my mind’s eye when listening to music make me feel energetic”) and the second one to soothing visual imagery (#10 “The images I see in my mind’s eye when listening to music make me feel relaxed”, #11 “The images I see in my mind’s eye when listening to music make me feel calm”).

<INSERT FIGURE 2 ABOUT HERE>

In essence, the revised analysis suggests that there is a general structure within the MVI questions that relates to vivid and soothing visual imagery, but this structure takes different forms across participants with different amounts of musical training. Whereas for both musically trained and untrained participants the vivid visual imagery is one of the main reasons why they listen to music, for musically trained participants these vivid images serve the purpose of increasing the arousal in the listener by making them feel more energetic and excited. On the other hand, for musically untrained individuals these vivid images are simply always present as an integral part of the listening experience but they do not seem to modulate the listeners' emotions explicitly. The soothing visual imagery is more similar across both groups. It makes both musically trained and untrained feel more relaxed and calm, while also enabling untrained individuals to dive into a different world and detach from everyday life.

To test whether vividness of visual imagery and the music-related visual imagery tap onto the same latent constructs, the scores from CFA analysis for each group were correlated with the VVIQ scores as well as the scores the five facets of musical sophistication index (Table 3). This analysis shows small albeit consistent positive correlations (we have reversed the original VVIQ scale to be more intuitive so that low scores in VVIQ indicate lower vividness) between the MVI factors, mainly between MVI factor Vivid Visual Imagery for musically untrained and trained (r=0.15 and r=0.29, respectively). This correlation indicates about 8% overlap between vividness of visual imagery and active musical visual imagery constructs, which is worth noting, but does not give reason to assume that the two measures are tapping onto the same ability. Musical training seems to influence the relationship between vividness of visual imagery and the active musical imagery. Although this suggests that the link has to do with expertise rather than differences in ability, the direction of this influence is beyond the present design.

Interestingly, most facets of the Gold-MSI and both MVI factors exhibit small correlations among the participants that possessed more formal musical training or exhibited otherwise moderate to high engagement with music. MVI factor SOOTHING VISUAL IMAGERY seems to be associated more strongly to Musical Training, Perceptual, and Active components in the Gold-MSI than with the VIVID VISUAL IMAGERY. The pattern between musical sophistication and MVI factors suggest that music-related visual imagery is related to expertise and particularly the active emotion regulation with imagery might show this relationship.

<INSERT TABLE 3 HERE>

The activities with other arts were only weakly or not at all correlated with the MVI factors scores (Active Arts – Fiction, $|r|<.13$, p=ns, Active Arts – Plays and theatre, $|r|<.10$, p=ns, but Active Arts – Visual Arts and Vivid Imagery $r=.24$, $p=.001$, although other factor, $r=0.02$, p=ns). The only statistically significant correlation between the VIVID VISUAL IMAGERY factor and activities in visual arts does make sense although the directionality of this relationship remains to be explored.

Content of music-related visual imagery. Asking their participants to rate the relevance of more than 600 expressions related to visual imagery and metaphors, Schaerlaeken et al. (same special issue) were able to show that the experience of visual imagery during music listening can be characterized by the 5 factors “Flow”, “Movement”, “Force”, “Interior”, and “Wandering”. While this study nicely delineates a potential structure underlying the (affective) experience of visual imagery during music listening, it does not provide the content or nature of the inner images themselves. We therefore asked our participants in an open-ended question to describe the images they conjure up during music listening in as much detail as possible. The most frequent type of visual image during music listening (33 out of 169) is a landscape or a scene from nature, followed by some autobiographical scene or event from the past (28) and images of people (28). Participants also often imagined a musical performance (21), including details of the performer and the performance venue. Another theme appearing frequently is an image of oneself (17), whether as a performer, member of the audience or character in a (fictive) narrative. Needless to say, the image of oneself being a world-class performer belongs, for the vast majority, just as much in an imaginary world as seeing oneself flying over a landscape or into outer space.

On the other hand, many people also experience more abstract visual imagery. Different colours and shades (20), animated shapes (20) and geometric objects and patterns (10) were frequently reported by our participants. Musically trained participants, perhaps not surprisingly, often see images related to the musical structure, such as melodic or instrument lines (5), the musical score (4), but also harmony (2) and tempo (1).

It is important to note that these emerging themes are not exclusive, i.e. individuals can experience a mix of concrete and abstract visual imagery, and several categories can be present at once or consecutively. What is more, some musical pieces may elicit consistently the same (autobiographical) image while the inner images evoked by other musical excerpts vary from one to another listening situation. These differences can occur both within as well as across participants. It should also be noted that the above-mentioned themes of music-related visual imagery occur across both VIVID VISUAL IMAGERY and SOOTHING VISUAL IMAGERY. Both types of visual imagery contain a multiplicity of concrete and abstract images that are shaped by highly personal experiences and associations.

In sum, it has become apparent that visual imagery during music listening is a highly flexible and idiosyncratic phenomenon. Consistent mappings between music and images over time might occur in special cases (i.e. synaesthesia), but are rather an exception. Much more common are free associations, whether concrete or abstract, that draw on experiences from everyday life. Most images are ephemeral, dynamic, evocative and/or reflective. They allow the mind to wander from one image to the next, zoom in and out of the music (or the images) and often seem to serve the purpose of modulating the emotional experience by decreasing or increasing the arousal.

Discussion

We offer the first study in which the content of visual imagery during music listening is systematically studied in a representative sample, employing a new set of independent questionnaire items to study this experience. Our first aim was to find out how common visual imagery is and to investigate the different kinds of inner images that individuals commonly experience when listening to music. We established that music-related visual imagery is frequently experienced; 77% of a representative sample indicated that they have seen images in their mind's eye during music listening. Using a combination of cluster and factor analyses, we were further able to demonstrate that music-related visual imagery consists of two components, VIVID VISUAL IMAGERY and SOOTHING VISUAL IMAGERY. The latter relates to emotional effects of imagery that usually pertain to relaxing and calm emotions. This component seems to be marginally influenced by musical training. The other component, VIVID VISUAL IMAGERY, was regarded to be one of the fundamental reasons for music listening by all participants. However, for those who are highly musically trained, VIVID VISUAL IMAGERY seems to produce energising experiences. For those who are less musically trained, this intimate connection to emotions simply was not relevant, although the vividness was still often mentioned as one of the main reasons for music listening.

Our second goal was to explore the extent to which visual imagery in response to music is different from visual imagery in general. While there were small correlations between the vividness of visual imagery ability and music-related visual imagery factors, the overlap between the two was negligible. We surmise that this suggests separate processes governing generic visual imagery and music-related visual imagery.

Our third aim was to investigate how visual imagery correlates with domain-specific skills such as musical training. Our observations that music-related visual imagery is influenced by musical training—whereas the vividness of visual imagery is little affected by training (Sunday et al., 2017)—indirectly supports this line of thought. The impact of musical training on visualizations of sound and music has been shown before (Küssner & Leech-Wilkinson, 2014). Whereas in that study musically trained individuals showed a smaller range of audio-visual mapping strategies than musically untrained participants, our results indicate that both trained and untrained respondents show a great diversity of visual imagery. However, the functional uses of visual imagery might differ dependent on the amount of musical training. Given that musically trained individuals seem to use inner images to modulate their level of arousal, either by increasing (VIVID VISUAL IMAGERY) or decreasing it (SOOTHING VISUAL IMAGERY), the differences between voluntary and involuntary visual imagery in a musical context need to be investigated further (see also

Taruffi & Küssner, same special issue). While some music might automatically trigger culturally mediated iconic imagery (e.g., a fanfare suggesting the image of a hunt) the evocation of certain images is of course highly mediated by learning. On the other hand, individuals might proactively conjure up certain images when listening to music to regulate their mood (Saarikallio, 2011; van Goethem & Sloboda, 2011).

Juslin and Västfjäll's (2008) hypothesized relation between visual imagery and felt emotions was one of the starting points of this study. Although the present study cannot answer the question whether visual imagery is causally linked to emotions felt during music listening, our results provide further evidence that visual imagery is intricately woven into emotional responses to music by modulating the listener's arousal. Recent constructionist account of emotional meanings in music would also be consistent with such close linking of emotions and imagery; it assumes that the information from different sensory domains contributes to the active meaning-making process (Cespedes-Guevara & Eerola, 2018). Others have argued that emotional responses and visual imagery are co-occurring, but separate phenomena which enable listeners to get absorbed by the music (Tellegen & Atkinson, 1974). Interestingly, Vroegh (same special issue) showed that emotions and visual imagery may have completely different interrelations depending on the listener's state of absorption. He distinguishes between 'zoning in' (low meta-awareness of state of mind, intense experience) and 'tuning in' (higher meta-awareness, less intense experience) states of absorption. During the latter state, visual imagery is most strongly coupled with positive affect and also with an altered state of consciousness. During a 'zoning in' experience, visual imagery is connected to (autobiographical) memories and attentional focus.

Visual imagery might also be regarded as a subcategory of musical mind-wandering. Both Taruffi et al. (2017) and Vroegh (same special issue) have shown that spontaneous thoughts during music listening occur most frequently in the shape of visual imagery rather than words. Visual imagery is also a key feature of musical daydreams (Herbert, 2011) and may be modulated by internally as well as externally oriented attention processes.

To conclude, we were able to demonstrate that visual imagery during music listening plays a pivotal role for the majority of listeners, independent of their musical background, and one of the main functions of visual imagery might be the (voluntary) modulation of emotional arousal, making people feel either more energetic or calm. However, many open questions remain, particularly with regards to the causal role in music-induced emotions. Despite the challenges of measuring objectively such a highly personal experience, studying music-related visual imagery offers, at the same time, the potential to discover new aspects of the multimodal listening experience.

References

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. (2000). Music training and mental imagery ability. *Neuropsychologia, 38*(12), 1664 - 1668.

Balteș, F. R., & Miu, A. C. (2014). Emotions during live music performance: Links with individual differences in empathy, visual imagery, and mood. *Psychomusicology: Music, Mind, and Brain, 24*(1), 58-65.

Band, J., Quilter, S., & Miller, G. (2001). The influence of selected music and inductions on mental imagery: Implications for practitioners of Guided Imagery and Music. *Journal of the Association for Music and Imagery, 8*, 13-33.

Boltz, M. G., Ebendorf, B., & Field, B. (2009). Audiovisual Interactions: the Impact of Visual Information on Music Perception and Memory. *Music Perception: An Interdisciplinary Journal, 27*(1), 43-59. doi: 10.1525/mp.2009.27.1.43

Bonde, L. O. (2005). "Finding a New Place…" Metaphor and Narrative in One Cancer Survivor's BMGIM Therapy. *Nordic Journal of Music Therapy, 14*(2), 137-154.

Bonde, L. O. (2015). The Bonny Method of Guided Imagery and Music (GIM) in Europe. *Approaches: Music Therapy & Special Music Education, 7*(1), 86-90.

Bonny, H. L. (2002). *Music & consciousness: The evolution of guided imagery and music*. Barcelona Publishers.

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition, 54*(2), 103-109.

Broggin, E., Savazzi, S., & Marzi, C. A. (2012). Similar effects of visual perception and imagery on simple reaction time. *The Quarterly Journal of Experimental Psychology, 65*(1), 151-164.

Bullerjahn, C. (2016). *Grundlagen der Wirkung von Filmmusik (3rd ed.)*. Augsburg: Wißner-Verlag.

Campos, A. (2011). Internal Consistency and Construct Validity of Two Versions of the Revised Vividness of Visual Imagery Questionnaire. *Perceptual and Motor Skills, 113*(2), 454-460. doi: 10.2466/04.22.pms.113.5.454-460.

Cespedes-Guevara, J. & Eerola, T. (2018). Music Communicates Affects, Not Basic Emotions - A Constructionist Account of Attribution of Emotional Meanings to Music. *Frontiers in Psychology, 9*, 215. https://www.frontiersin.org/article/10.3389/fpsyg.2018.00215

Cohen, A. J. (2011). Music as a source of emotion in film. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications.* Oxford: Oxford University Press.

Day, R. A., & Thompson, W. F. (same special issue). Measuring the onset of experiences of emotion and imagery in response to music. *Psychomusicology: Music, Mind, and Brain.*

Eitan, Z. (2017). Musical Connections: Crossmodal Correspondences. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition* (pp. 213-224). New York: Taylor & Francis.

Eldar, E., Ganor, O., Admon, R., Bleich, A., & Hendler, T. (2007). Feeling the Real World: Limbic Response to Music Depends on Related Content. *Cerebral Cortex, 17*(12), 2828-2840. doi: 10.1093/cercor/bhm011

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods,* 4(3), 272-299.

Geringer, J. M., Cassidy, J. W., & Byo, J. L. (1997). Nonmusic Majors' Cognitive and Affective Responses to Performance and Programmatic Music Videos. *Journal of Research in Music Education, 45*(2), 221-233. doi: 10.2307/3345582

Goldberg, M. E., Chattopadhyay, A., Gorn, G. J., & Rosenblatt, J. (1993). Music, music videos, and wear out. *Psychology and Marketing, 10*(1), 1-13. doi: 10.1002/mar.4220100102

Herbert, R. (2011). *Everyday Listening: Absorption, Dissociation and Trancing*. London: Routledge.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews, 10*(3), 235-266. doi: 10.1016/j.plrev.2013.05.008

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion, 8*(5), 668-683.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31*(5), 559-575.

Karagozoglu, S., Tekyasar, F., & Yilmaz, F. A. (2013). Effects of music therapy and guided visual imagery on chemotherapy-induced anxiety and nausea–vomiting. *Journal of Clinical Nursing, 22*(1-2), 39-50.

Kline, R. B. (2010). *Principles and practice of structural equation modeling (3rd ed.)*. New York, New York: Guilford Press.

Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., & Jacobs, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *NeuroImage, 81*(1), 49-60.

Kolers, P. A. (1983). Perception and representation. *Annual Review of Psychology, 34*(1), 129-166.

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J., Ganis, G., . . . Alpert, N. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science, 284*(5411), 167-170.

Kuchinke, L., Kappelhoff, H., & Koelsch, S. (2013). Emotion and music in narrative films: A neuroscientific perspective. In S.-L. Tan, A. J. Cohen, S. D. Lipscomb & R. A. Kendall (Eds.), *The Psychology of Music in Multimedia*. New York: Oxford University Press.

Küssner, M. B. (2013). Music and shape. *Literary and Linguistic Computing, 28*(3), 472-479. doi: 10.1093/llc/fqs071

Küssner, M. B. & Leech-Wilkinson, D. (2014). Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing paradigm. *Psychology of Music, 42*(3), 448-469. doi: 10.1177/0305735613482022

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*(1), 17-24.

Müllensiefen, D., Gingras, B., Musil, J., Stewart, L. (2014) The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE 9*(2)*: e89642*. https://doi.org/10.1371/journal.pone.0089642

Osborne, J. W. (1981). The mapping of thoughts, emotions, sensations, and images as responses to music. *Journal of Mental Imagery, 5*, 133-136.

Quittner, A., & Glueckauf, R. (1983). The facilitative effects of music on visual imagery: A multiple measures approach. *Journal of Mental Imagery, 7*, 105-120.

Revelle, W., & Rocklin, T. (1979). Very simple structure-alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research, 14*(4), 403-414.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65.

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of a known factorial structure. *Psychological Assessment, 24*(2), 282–292.

Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychology of Music, 39*(3), 307-327. doi: 10.1177/0305735610374894

Schaerlaeken, S., Glowinski, D., Rappaz, M.-A., & Grandjean, D. (same special issue). Visual Imagery and Metaphors Evoked by the Sound Listening of Classical Music: Characterization, Classification, and Measurement. *Psychomusicology: Music, Mind, and Brain.*

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99(6)*, 323-338.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971-995. doi: 10.3758/s13414-010-0073-7

Stratton, V. N., & Zalanowski, A. H. (1992). The Interfering Effects of Music with Imagery. *Imagination, Cognition and Personality, 11*(4), 381-388. doi: 10.2190/wa5e-cr32-q549-mx77

Sunday, M., McGugin, R. W., & Gauthier, I. (2017). Domain-specific reports of visual imagery vividness are not related to perceptual expertise. *Behavior Research Methods, 49*(2), 733-738. doi: 10.3758/s13428-016-0730-4

Tan, S.-L., & Kelly, M. E. (2004). Graphic representations of short musical compositions. *Psychology of Music, 32*(2), 191-212. doi: 10.1177/0305735604041494

Taruffi, L., & Küssner, M. B. (same special issue). A Review of Music-Evoked Visual Mental Imagery: Conceptual Issues, Relation to Emotion, and Functional Outcome. *Psychomusicology: Music, Mind, and Brain.*

Taruffi, L., Pehrs, C., Skouras, S., & Koelsch, S. (2017). Effects of Sad and Happy Music on Mind-Wandering and the Default Mode Network. *Scientific Reports, 7*(1), 14396. doi: 10.1038/s41598-017-14849-0

Tellegen, A., & Atkinson, G. (1974). Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology, 83*(3), 268-277. doi: 10.1037/h0036681

van Goethem, A., & Sloboda, J. (2011). The functions of music for affect regulation. *Musicae Scientiae, 15*(2), 208-228. doi: 10.1177/1029864911401174

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*(1), 80-113. doi: 10.1016/j.cognition.2005.09.003

Vroegh, T. (same special issue). Zoning in or Tuning in? Identifying Distinct Absorption States in Response to Music. *Psychomusicology: Music, Mind, and Brain.*

Vuoskoski, J. K., & Eerola, T. (2015). Extramusical information contributes to emotions induced by music. *Psychology of Music, 43*(2), 262-274. doi: 10.1177/0305735613502373

Ximénez, C. (2009). Recovery of weak factor loadings in confirmatory factor analysis under conditions of model misspecification. *Behavior Research Methods, 41*(4), 1038-1052.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

Tables

Table 1

Table 1. Loadings from the EFA and CFA. Loadings under <.450 are not displayed.

Item	**F1 EFA/CFA**	**F2 EFA/CFA**	**F3 EFA/CFA**
#1 I often conjure up images while listening to recorded music	.71/.76		
#2 I often conjure up images while listening to live music	.46/NA		
#3 I rarely see images in my mind's eye when listening to live or recorded music	-.61/NA		
#8 I try to manipulate internal images that come up while I listen to music	.47/NA		
#12 The images I see in my mind's eye when listening to music make me feel excited	.50/.70		
#14 When I listen to music I see abstract figures and shapes in my mind's eye	.53/NA		
#17 The images I see in my mind's eye when listening to music are dynamic	.64/NA		
#18 I see images in my mind's eye whenever I listen to music	.58/.66		
#23 I often conjure up images while listening to music with eyes open	.68/.63		
#24 I often conjure up images while listening to music with eyes closed	.49/NA		
#10 The images I see in my mind's eye when listening to music make me feel relaxed		.81/.89	
#11 The images I see in my mind's eye when listening to music make me feel calm		.77/.90	
#7 I try to suppress internal images that come up while I listen to music			.69/.64
#15 I am often bothered by the images I see in my mind's eye when listening to music			.69/.75
#16 The images I see in my mind's eye when listening to music are static			.45/.52
Loadings	4.78	2.28	1.50
Variance explained	21%	10%	7%

Note. NA refers to items deleted in the CFA (loadings <.50).

Table 2

Table 2. The fit indices of the CFA models.

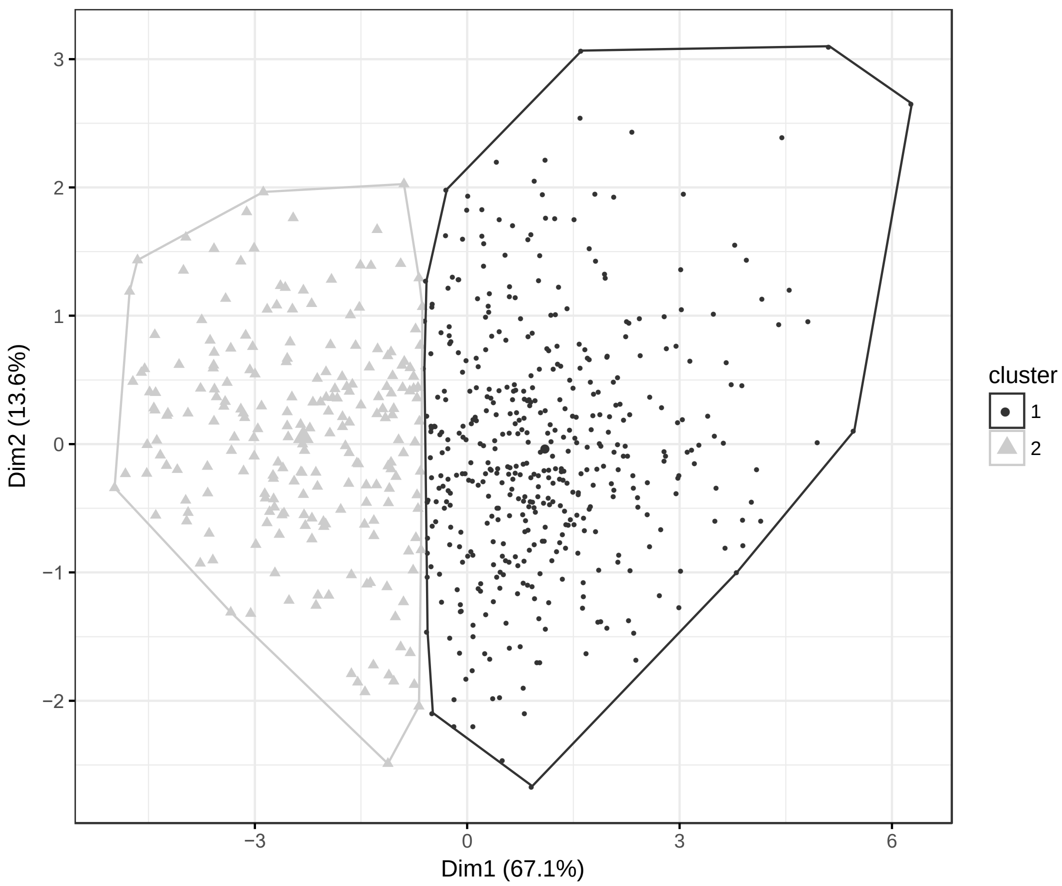
Model	X^2	*df*	CFI	RMSEA	RMSEA CI_{90}
Overall model (n=669) 3 factors	68.88	24	0.945	0.084	0.061-0.107
Musically untrained (n=450) 2 FA	10.01	8	0.989	0.037	0.001-0.100
Musically trained (n=219) 2 FA	7.07	4	0.988	0.094	0.001-0.205

Table 3. Correlations between the MVI factor scores, VVIQ and musical training across the two clusters representing musically trained and untrained participants (the grouping from the cluster analysis).

	Musical Training	
	Untrained	Trained
	VVIQ	
MVI Vivid	0.15**	0.29***
MVI Soothing	0.12*	0.18**
	MSI Training	
MVI Vivid	-0.06	-0.34***
MVI Soothing	-0.06	-0.44***
	MSI Perceptual	
MVI Vivid	0.08	-0.18**
MVI Soothing	0.02	-0.28***
	MSI Emotional	
MVI Vivid	0.11	0.28***
MVI Soothing	0.11	0.23***
	MSI Singing	
MVI Vivid	0.03	-0.06
MVI Soothing	-0.08	-0.08
	MSI Active	
MVI Vivid	0.22**	-0.13*
MVI Soothing	0.16*	-0.19**

*** p<.001, ** p<.01, * p<.05.

Figures


cluster
1
2
Dim2 (13.6%)
Dim1 (67.1%)
3
2
1
0
−1
−2
−3
0
3
6

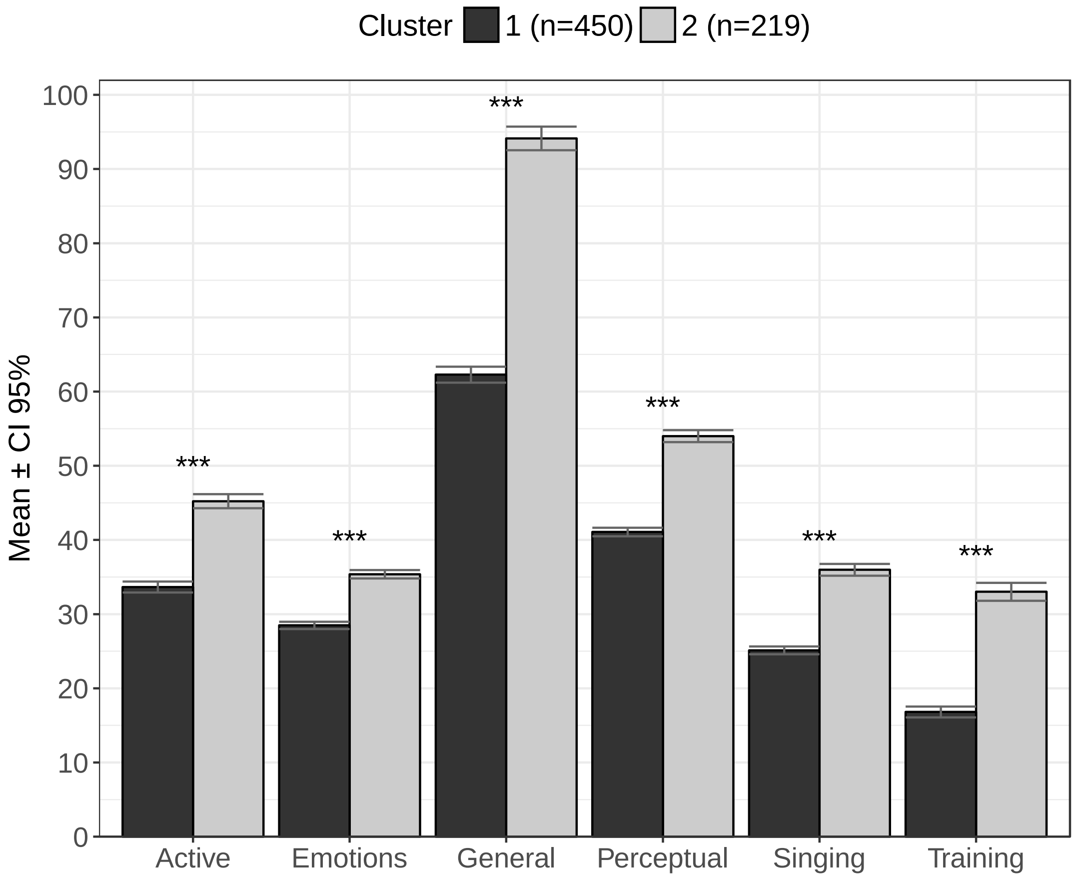

Cluster 1 (n=450) 2 (n=219)
Mean ± CI 95%
***
Active
Emotions
General
Perceptual
Singing
Training
0
10
20
30
40
50
60
70
80
90
100

Figure 1. Descriptive summary of the two clusters based on Gold-MSI variables.

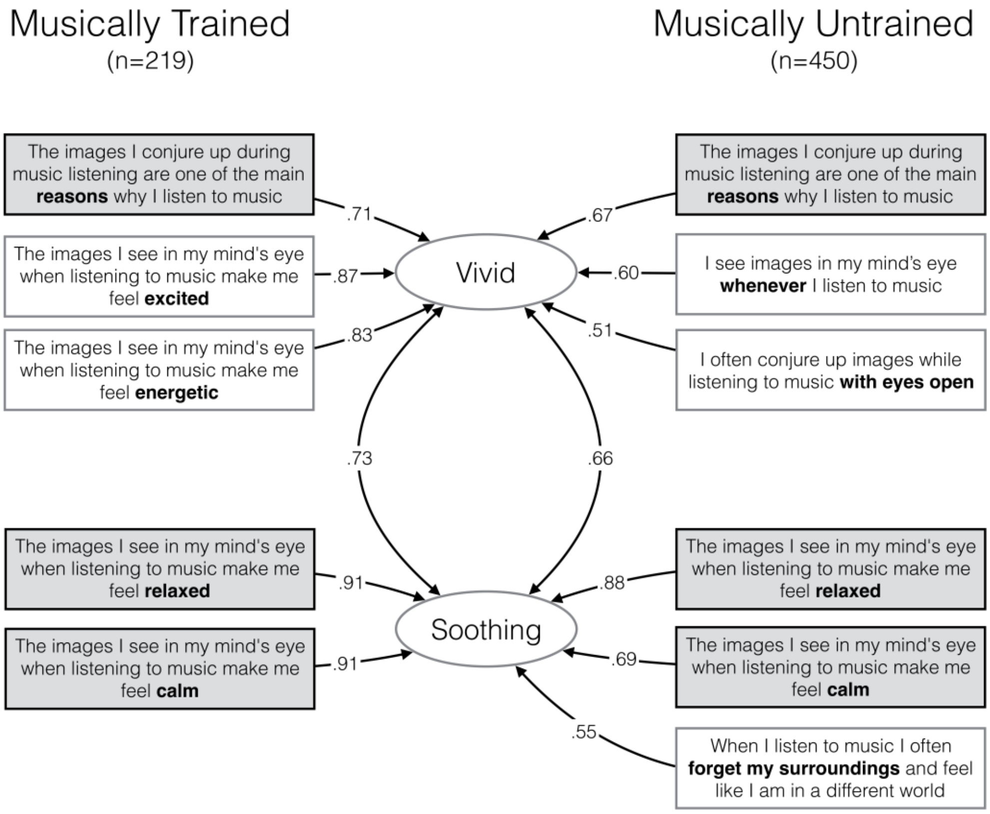


Figure 2. CFA models for musically trained and untrained participants.

Appendix A

24 items of the Music-Related Visual Imagery (MVI) instrument and the mean ratings.

Item	Mean (SD)
#1 I often conjure up images while listening to recorded music.	5.08 (1.27)
#2 I often conjure up images while listening to live music.	4.02 (1.32)
#3 I rarely see images in my mind’s eye when listening to live or recorded music.	3.19 (1.46)
#4 When I read leaflets or programme notes prior to a musical performance I often conjure up images during the concert.	3.86 (1.34)
#5 The images I conjure up during music listening are one of the main reasons why I listen to music.	3.79 (1.43)
#6 I listen to music because it allows me to immerse myself into a different world.	5.20 (1.27)
#7 I try to suppress internal images that come up while I listen to music.	2.99 (1.31)
#8 I try to manipulate internal images that come up while I listen to music.	4.01 (1.32)
#9 When I listen to music I often forget my surroundings and feel like I am in a different world.	5.05 (1.22)
#10 The images I see in my mind’s eye when listening to music make me feel relaxed.	4.94 (1.14)
#11 The images I see in my mind’s eye when listening to music make me feel calm.	4.89 (1.10)
#12 The images I see in my mind’s eye when listening to music make me feel excited.	4.76 (1.18)
#13 The images I see in my mind’s eye when listening to music make me	4.64 (1.24)

feel energetic.	
#14 When I listen to music I see abstract figures and shapes in my mind’s eye.	3.87 (1.37)
#15 I am often bothered by the images I see in my mind’s eye when listening to music.	2.77 (1.27)
#16 The images I see in my mind’s eye when listening to music are static.	3.16 (1.17)
#17 The images I see in my mind’s eye when listening to music are dynamic.	4.70 (1.19)
#18 I see images in my mind’s eye whenever I listen to music.	4.22 (1.36)
#19 The images in my mind's eye during music listening occur spontaneously.	4.87 (1.11)
#20 The images I conjure up during music listening only last a few moments.	3.99 (1.18)
#21 The images I conjure up during music listening are often substituted by new images.	4.21 (1.18)
#22 I often conjure up concrete scenes (e.g. landscapes, people, etc).	4.57 (1.39)
#23 I often conjure up images while listening to music with eyes open.	4.63 (1.37)
#24 I often conjure up images while listening to music with eyes closed.	4.80 (1.35)